A proline-rich transcriptional activation domain in murine HOXD-4 (HOX-4.2)

Isabel Rambaldi[1], Erzsebet Nagy Kovàcs[1] and Mark S.Featherstone[1,2,*]

[1]McGill Cancer Centre and [2]Departments of Oncology and Medicine, McGill University, 3655 Drummond Street, Montréal, Québec H3G 1Y6, Canada

ABSTRACT

The product of the murine *Hoxd-4* (*Hox-4.2*) gene is a transcription factor that acts upon an autoregulatory element in *Hoxd-4* upstream sequences (1). Using this activity as an assay in transient transfections of P19 embryonal carcinoma (EC) cells, we performed a mutational analysis to map functional domains in the HOXD-4 protein. The importance of the homeodomain was shown by a single amino acid change in this region that abolished activity. Deletion analysis revealed that many evolutionarily conserved regions outside of the homeodomain were dispensable for activation in our assay. Fusions to the GAL4 DNA-binding domain mapped a transcriptional activation function to the HOXD-4 proline-rich N-terminus. The proline-rich transcription factor AP2 squelched activation by HOXD-4 and by GAL4/HOXD-4 N-terminus fusion proteins. Together, these results suggest that HOXD-4 harbors a transcriptional activation domain of the proline-rich type.

INTRODUCTION

The homeotic genes of insects and the *Hox* genes of vertebrates are evolutionarily related families represented over a broad range of animal species (2). Homeotic and *Hox* genes exert their effects through the transcriptional regulation of target genes, few of which have been identified. The best characterized example is the *Drosophila* homeotic gene *Deformed* (*Dfd*). The Dfd protein has been shown to positively regulate the expression of its own gene (3). A minimal *Dfd* autoregulatory enhancer has four high affinity Dfd binding sites (4).

DNA-binding is mediated by the 61 residue homeodomain encoded by the highly conserved homeobox (5–7) and found in the C-terminal half of the protein. Apart from the homeodomain, other regions of the *Hox* and homeotic gene products are also conserved. Most notable is the so-called conserved pentapeptide (8, 9), a domain found just N-terminal to the homeodomain of many Antp class homeoproteins. Regions of some conservation are also found immediately C-terminal to the homeodomain and at the extreme N-terminus (2). The extent of homology is greatest amongst members of the same subfamily.

Other regions can be identified based on amino acid content or character, rather than percent identity. For example, homeoproteins of the Dfd subfamily have a proline-rich region in the N-terminal half of the molecule (10–12). The C-terminus of the mouse HOXD-4 protein is very serine-rich (10). Determining the roles of these regions is an important approach to understanding differential function among the *Hox* gene products.

Via transient transfections of P19 embryonal carcinoma (EC) cells, we have recently identified an autoregulatory element of the murine *Hoxd-4* gene (1). This activity provided us with an assay by which to assess the functional contributions of regions within the HOXD-4 homeoprotein to transcriptional activation. In a complementary test, HOXD-4 residues were fused to the GAL4 DNA-binding domain to map transcriptional activation functions. Using these approaches, we have shown that many regions of HOXD-4 are dispensable for activation of the *Hoxd-4* promoter in our assay. The proline-rich N-terminus of HOXD-4 is a strong transcriptional activation domain.

MATERIALS AND METHODS

Plasmids

We use the recently adopted *Hox* nomenclature (13). Old *Hox* names are given in brackets on first use. Two sequencing errors and an error in the *Hoxd-4* cDNA were included in the previously published sequence of *Hoxd-4* (10). The corrected sequence results in the following changes: G39 to S39, R95 to P95, and P192 to S192 (Fig. 1). Note that with S192 the sequence of the HOXD-4 homeodomain now shares 100% identity with that of human HOXD4 (9). The sequence of *Hoxd-4* in the GenBank database has been revised appropriately (accession number J03770).

Hoxd-4 expression vectors were based on the murine *phosphoglycerate kinase-1* (*pgk-1*) promoter and 3′ processing sequences (14) which were a kind gift of Dr Michael McBurney. *Hoxd-4* coding region sequences were obtained from cDNA and genomic clones (10) and inserted between the *pgk-1* elements in pBluescribe (Stratagene). The unique *Xho* I, *Bam* HI, and *Xba* I sites (Fig. 2) were engineered via PCR-mediated mutagenesis and standard recombinant DNA techniques.

---

* To whom correspondence should be addressed at: McGill Cancer Center, Room 714, 3655 Drummond Street, Montréal, Quebec H3G 1Y6, Canada

Sequences naturally flanking the ATG of the third codon conform to a Kozak consensus (15) whereas those for the first codon do not (10), suggesting that translation of *Hoxd-4* transcripts begins with the methionine at position 3 (Fig. 1). Creation of the *Bam* HI site involved restriction at a *Bal* I site and removed approximately 300bp of leader sequence and the upstream initiation codon (M1) (Fig. 1). The Kozak consensus at the initiation codon for M3 was lost in this process. A Kozak consensus sequence was restored by site-directed mutagenesis to give the expression vector p4.2 (Table 1). p4.2-transfected cells synthesize a product recognized by polyclonal antibodies directed against the HOXD-4 N- and C-termini (Fig. 3). The apparent size of this protein on Western blots is larger than expected, also noted for the products of the chicken *Hoxd-4* and *Hoxa-4* homologs which are also proline-rich (16).
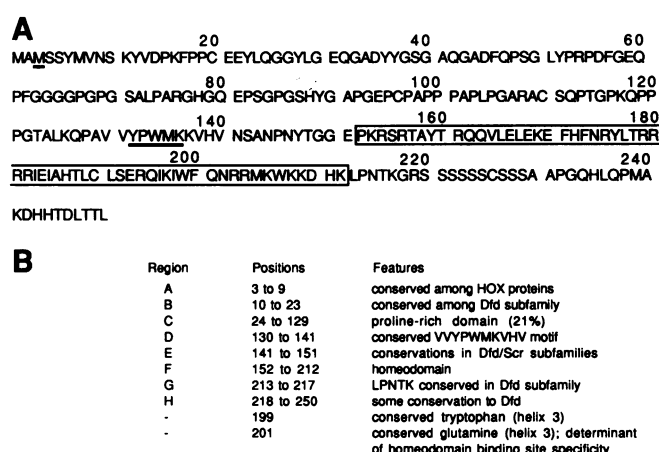
A

```
         20                40                60
MAMSSYMVNS KYVDPKFPPC EEYLQGGYLG EQGADYYGSG AQGADFQPSG LYPRPDFGEQ
         80               100               120
PFGGGGPGPG SALPARGHGQ EPSGPGSHYG APGEPCPAPP PAPLPGARAC SQPTGPKQPP
        140               160               180
PGTALKQPAV VYPWMKKVHV NSANPNYTGG E PKRSRTAYT RQQVLELEKE FHFNRYLTRR
        200               220               240
RRIEIAHTLC LSERQIKIWF QNRRMKWKKD HK LPNTKGRS SSSSSCSSSA APGQHLQPMA

KDHHTDLTTL
```

B

Region	Positions	Features
A	3 to 9	conserved among HOX proteins
B	10 to 23	conserved among Dfd subfamily
C	24 to 129	proline-rich domain (21%)
D	130 to 141	conserved VVYPWMKVHV motif
E	141 to 151	conservations in Dfd/Scr subfamilies
F	152 to 212	homeodomain
G	213 to 217	LPNTK conserved in Dfd subfamily
H	218 to 250	some conservation to Dfd
-	199	conserved tryptophan (helix 3)
-	201	conserved glutamine (helix 3); determinant of homeodomain binding site specificity

Figure 1. A. Primary amino acid sequence of HOXD-4 given in one-letter code. Constructs used in these studies are derived from p4.2 and express a HOXD-4 protein starting with the second methionine at position 3 (underlined). The homeodomain is boxed, and the conserved pentapeptide underlined. Corrections to the previously published sequence of HOXD-4 (10) are noted in Materials and Methods. **B.** Features of HOXD-4 used to guide a mutagenesis strategy. Numbering as in A above.

Derivatives of p4.2 mutated in the *Hoxd-4* coding region were generated by oligonucleotide-directed mutagenesis on single-stranded DNA (deletions and substitutions) and by standard recombinant DNA techniques (Table 1). The procedure of Kunkel *et al.* (17) was used for site-directed mutagenesis. All mutants were verified by sequencing of double-stranded DNA by the chain termination method (18). Mutant vectors gave products of appropriate size in Western blot analysis.

The construction of the luciferase reporter pSXΔP2 under the control of the *Hoxd-4* promoter/autoregulatory element is described elsewhere (1). An expression vector for the GAL4 DNA binding domain (residues 1 to 148) was made from the vector pSG5Gal-VP16 (19) (kind gift of Dr John White) by enzymatic removal of VP16 coding sequences. *Hoxd-4* coding sequences were introduced in frame with the *Gal4* coding region. The GAL4-responsive luciferase reporter pXPAL7 was made by PCR amplification of the promoter region of the vector pAL7 (20) (kind gift of Dr John White) and transfer to the luciferase vector pXP1 (21). Luciferase reporter p34a was made by H.Pöpperl (unpublished). All constructs were verified by sequence analysis. The AP2 expression vector sp(RSV)AP2 was a kind gift of Dr Trevor Williams (22).

Cell culture, transfections and enzymatic assays

P19 EC cells were cultured and calcium phosphate transfection assays performed as described previously (1). Transfections of differentiating cells were done 48 hours after retinoic acid (RA) addition to 3×10^{-7}M. For assays on the *Hoxd-4* promoter, each dish received 1μg of expression vector, 4μg of luciferase reporter pSXΔP2, and 1μg of pRSVlacZ as internal standard. For assays with GAL4 fusions, the amount of expression vector was reduced to 100ng. Precipitates for each 100mm dish were brought to 10μg total DNA with supercoiled pBluescribe (Stratagene). All luciferase values were normalized to those for β-galactosidase.

Antibody production and Western analysis

Fusions of the N-terminal half or C-terminal half of HOXD-4 to the amino terminus of trpE were accomplished through the pATH2 and pATH3 vectors, respectively (23). Immunizations were according to Harlow and Lane (24). Approximately 125μg

Table 1. Mutations in the *Hoxd-4* coding region

Vector	Manipulation	Nucleotide sequence across site of mutation[a]	Protein sequence across junction[a,b]	Residues altered[c]
p4.2	deletion, insertion	J: GACCTGCAAGGATCCA/CCATGAGTTCGTAT	3 /MSSYM 7	1 and 2
pdlB	deletion	O: TTCGTATATATGGTGAAC/TTGCAGGGTGGCTACC	3 MSSYMVN/LQGGY 28	10 to 23
pdlAB	deletion	O: GCAAGGATCCACCATG/TTGCAGGGTGGCTACC	3 M/LQGGY 28	4 to 23
pdlD	*Acc* I/*Eco* RI deletion and adaptor insertion	O: CT/TC/CACGTG O: AATTCACGTG/GA/	128 PAVV/f/HVNS 142	132 to 138
pdlG	deletion	O: GAAAAAAGACCACAAA/GGCAGGTCTTCTTCCT	209 KDHK/GRSS 221	213 to 217
pdlGH	deletion	O: GAAAAAAGACCACAAA/TAGAAGTGGGGACCTT	209 KDHK/x	213 to 250
psbK201	substitution	O: CAAGATCTGGTTC/A/AGAACCGGAGGA	197 KIWF/k/NRRM 205	201
psbY199	substitution	J: GATCAAGATCT/AT/TTCCAGAACCGG	195 QIKI/y/FQNR 203	199
p4.2FS	*Bgl* II, blunt, ligate	J: CAGATCAAGATC/GATCTGGTT	195 QIKI/dlv...	199 onward

[a]'J', for junction, indicates the sequence across the site of mutation. 'O' designates the actual oligonucleotide used in the mutagenesis. Slashes indicate boundaries between regions juxtaposed or created as a result of the manipulations.

[b]Small letters indicate amino acids not found in the wild-type HOXD-4 protein. 'x' indicates the presence of a termination codon in the corresponding transcript. Numbering as in figure 1.

[c]Includes deletions and substitutions. Numbering of altered residues is inclusive.

of each induced fusion protein were electroeluted from SDS/polyacrylamide gels, lyophilized, resuspended in 500μl water and emulsified with 500μl of complete Freund's adjuvant. Rabbits were injected subcutaneously at 4 sites each, 250μl per site. Boosts were given at 4 week intervals using incomplete Freund's adjuvant. Serum was obtained at the same time by ear bleeding (25ml) or by cardiac puncture 4 weeks after the third boost. Western blots were performed with Biorad minigel and minitransfer modules and a Biorad immunoblotting kit. A 1:100 dilution of the primary antibody was used with alkaline phosphatase-conjugated goat antirabbit second antibody. The availability of antisera independently directed against the HOXD-4 N- and C-termini allowed us to assess the levels of proteins mutated in one domain by using antiserum directed against the other, thus avoiding detection problems due to epitope loss.

RESULTS

The *Hoxd-4* coding region

The HOXD-4 protein can be divided into 8 regions based on homologies within and beyond the Dfd subfamily of proteins, and on the basis of amino acid composition (Fig. 1). We used these to guide a mutational analysis of HOXD-4 structure and function (Table 1). To assay HOXD-4 activity, we relied on its ability to activate transcription through an autoregulatory element in the *Hoxd-4* promoter in transient transfections of P19 EC cells (1). The wild-type expression vector for HOXD-4 (p4.2) and the

HOXD-4-responsive luciferase reporter under the control of the *Hoxd-4* promoter (pSXΔP2) are diagrammed in Figure 2.

After transfection into pluripotent and RA-treated P19 cells, p4.2 directed the synthesis of a product reactive on Western blots with antibodies against HOXD-4 (Fig. 3). Mutated derivatives of p4.2 were also examined for the production of immunoreactive species which were detected in all but four cases (Fig. 4). One of these four was p4.2FS which had a frame shift mutation in the homeobox in the region coding for the recognition helix (Table 1; Fig. 2). Transcriptional activation by p4.2 and its mutated derivatives was expressed as the ratio of standardized luciferase activity over that obtained with p4.2FS.

Transcriptional activation by mutant HOXD-4 proteins

Site-specific binding by the homeodomain is in part mediated by the third α helix (recognition helix) which makes contacts in the major groove (25−32). A key determinant of specificity is the residue at position 50 of the homeodomain, within the recognition helix (33−35). Substitution for lysine converts the specificity of binding to that of the bicoid homeodomain (30). As for all *Hox* and homeotic gene products, HOXD-4 has a glutamine at this position (Q201, Fig. 2). Conversion of this glutamine to lysine abrogated activation of the *Hoxd-4* promoter (sbK201, Fig. 4) and abolished binding to a HOXD-4 binding site in an electrophoretic mobility shift assay (data not shown). Nonetheless, this protein gave a band intensity comparable to wild-type on Western blots. This result underscores the importance of the homeodomain in transcriptional activation by HOXD-4. The tryptophan at position 199 of HOXD-4 is highly conserved and has been implicated in homeodomain function by genetic and structural observations. Conversion to a tyrosine residue (sbY199, Fig. 4) resulted in the loss of a detectable product on Western blots and a severe reduction in transcriptional activation.

Deletion of conserved N-terminal regions B, or A plus B (dlB and dlAB, Fig. 1 and 4), did not greatly affect HOXD-4 function, though the activity of dlAB was somewhat elevated. Products bearing deletions of much of the proline-rich region C or the
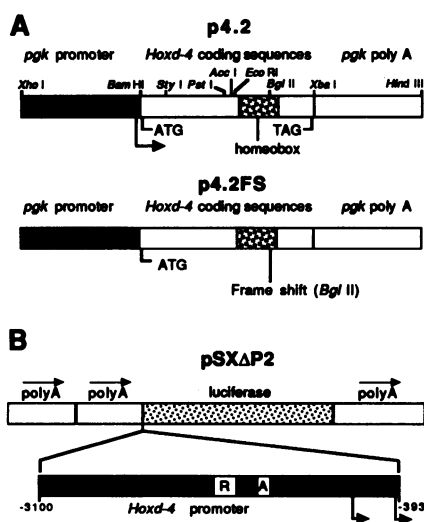

Figure 2. A. HOXD-4 expression vectors. Top: As indicated, the expression vector p4.2 contains *Hoxd-4* coding sequences flanked by the murine *pgk-1* (14) promoter and polyadenylation signal in pBluescribe. Direction of transcription, homeobox, start and stop codons and various restriction enzyme sites are given. Bottom: The control expression vector p4.2FS contains a frame-shift mutation at the *Bgl*II site in the homeobox. Not to scale. **B.** Luciferase reporter pSXDP2 is based on the vector pXP2 (21). It consists of tandem SV40 polyadenylation signals upstream of the firefly luciferase coding region (51) flanked 3' by a single SV40 polyadenylation signal. *Hoxd-4* promoter sequences (−393 to −3100 upstream of the start codon) are located in the multi-cloning site just 5' to the luciferase gene. Presumptive transcriptional start sites (10, 52) are given by subtended arrows. R and A indicate the presence of a retinoic acid response element (53) and an autoregulatory element (1), respectively. Not to scale.
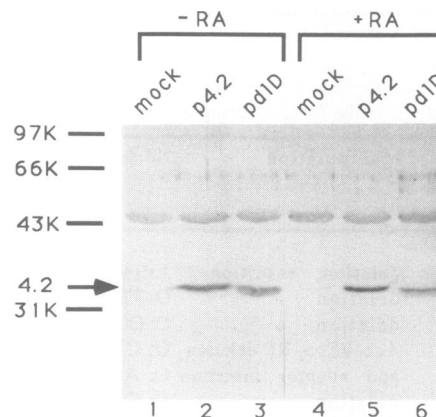

Figure 3. Western blot detection of HOXD-4 proteins in extracts of transfected cells. Extracts were prepared from pluripotent (−RA) and differentiating (+RA) P19 EC cells which had been mock-transfected (mock) or transfected with 4μg/plate of p4.2 or pdlD expression vectors (Fig. 4). 50μg of protein were loaded per lane. A *lacZ* expression vector was included to monitor transfection efficiency which was 1.8 fold higher in lane 2 than lane 3, and 1.2 fold higher in lane 5 than lane 6. Size markers in kilodaltons, as well as the position of HOXD-4 protein (arrow), are given at left.

	A B C DE F G H	FOLD ACTIVATION −RA		+RA		WESTERN
HOXD-4	WILD TYPE	14	(10)	14	(11)	+
dlB		15 ± 2 (3)		14 ± 3 (5)		+
dlAB		24 ± 7 (3)		15 ± 3 (5)		+
dlD		37 ± 5 (5)		22 ± 7 (9)		+
dlG		24 ± 1 (2)		12 ± 5 (5)		+
dlGH		7 ± 2 (5)		3 ± 2 (3)		+
sbHA-4	A-4	12 ± 0 (2)		11 ± 1 (2)		n.d.
sbK201	Q201→K201	2 ± 1 (6)		1 ± 0 (3)		+
sbY199	W199→Y199	3 ± 2 (3)		4 ± 1 (2)		-
HOXD4-FS	FRAME SHIFT	1 ± 1 (10)		1 ± 1 (11)		-

Figure 4. Effect of coding region mutations on transcriptional activation by HOXD-4 in pluripotent (−RA) and RA-treated (+RA) P19 EC cells. At top, representation of the wild-type HOXD-4 product of the p4.2 expression vector, divided into regions and lettered according to Figure 2. The homeodomain (region F) is stippled. Along left, names of mutants indicating deletion (dl), substitution (sb) or other alteration. 'A-4' designates the C-terminus of the *Hoxa-4* product. Single amino acid substitutions within the homeodomain at positions 199 and 201 are indicated. HOXD-4FS denotes the frame-shifted product of the p4.2FS expression vector. 'FOLD ACTIVATION' gives the ratio of luciferase activity in co-transfections with vectors encoding given HOXD-4 proteins over that obtained with the control vector p4.2FS. Activation by HOXD-4 was set to 14 (the average over multiple experiments) and other values normalized accordingly. Standard deviations are shown. The number of independent experiments for a given construct is given in brackets. '+' under 'WESTERN' indicates that a given protein could be detected at wild-type levels on Western blots of transfected cell extracts using rabbit polyclonal serum directed against the HOXD-4 N- or C-terminus (see Materials and Methods). 'n.d.' not determined.

entire N-terminal half of HOXD-4 could not be detected in Western analysis. A six amino acid deletion which removed four of the five amino acids of the conserved pentapeptide (8, 9) (dlD, Fig. 1 and 4) was 2.6 fold more active for transcriptional activation in pluripotent P19 cells. A five amino acid deletion of the conserved region G immediately C-terminal to the homeodomain (dlG, Fig. 1 and 4) also caused a modest increase in HOXD-4 activity. Full deletion of the C-terminus, including the serine-rich region H (dlGH, Fig. 4), impaired the ability of HOXD-4 to activate the *Hoxd-4* promoter. Replacement with the C-terminus of the paralogous *Hoxa-4* (*Hox-1.4*) product (sbHA-4, Fig. 4) restored activation to wild-type levels. Results were similar in both pluripotent and RA-treated P19 cells.

To ensure that transcriptional activation by HOXD-4 derivatives was mediated by the autoregulatory enhancer in the *Hoxd-4* promoter, we tested three HOXD-4 vectors on two different promoters. One was a minimal adenovirus major late promoter, while the other had, in addition, the *Hoxd-4* autoregulatory enhancer (1). As shown in Table 2, wild-type HOXD-4 as well as derivatives dlD and dlG were considerably more active on the promoter containing the autoregulatory element (pNBML). Consistent with results on the *Hoxd-4* promoter (Fig. 4), dlD is more active than wild-type. We conclude that the effects of HOXD-4 derivatives are mediated through the *Hoxd-4* autoregulatory enhancer.

Table 2. Transcriptional activation by HOXD-4 derivatives is dependent on the autoregulatory element in the *Hoxd-4* promoter[a]

Promoter	Fold activation[b] pML	pNBML	Specific increase mediated by the autoregulatory enhancer[c]
HOXD-4 vector			
none	1	1	1
p4.2	2	13	7
pdlD	4	27	7
pdlG	3	16	5

[a]HOXD-4 and two derivatives, dlD and dlG, were tested for specific activation through the autoregulatory enhancer in the *Hoxd-4* promoter in RA-treated P19 cells. Appropriate expression vectors were tested at 1 μg/100mm plate in cotransfections with luciferase reporters pML and pNBML. pML contains the adenovirus major late promoter. In pNBML, the 217bp *NruI/BamHI* fragment containing the *Hoxd-4* autoregulatory element has been inserted just upstream of the adenoviral TATA box (1).
[b]The fold increase over basal promoter activity mediated by HOXD-4 expression vectors is shown.
[c]The ratio of fold activation obtained with pNBML over pML.

A pXPAL7 / 5 x 17MER / TATA / LUCIFERASE

B

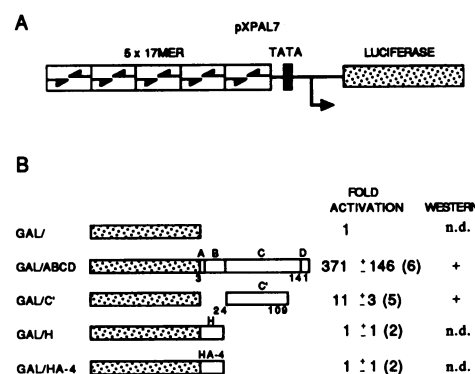|		FOLD ACTIVATION		WESTERN
GAL/		1		n.d.
GAL/ABCD	A B C D	371 ± 146 (6)		+
GAL/C'	C'	11 ± 3 (5)		+
GAL/H		1 ± 1 (2)		n.d.
GAL/HA-4	HA-4	1 ± 1 (2)		n.d.

Figure 5. Mapping of transcriptional activation domains in HOXD-4 by fusion to the GAL4 DNA binding domain. **A.** Schematic of luciferase reporter used to measure transcriptional activation by GAL4 derivatives. The promoter (20) consists of the adenovirus-2 major late TATA box and transcriptional start site (subtended arrow) downstream of five copies of the 17-mer GAL4 dimer binding site (boxed open arrows). **B.** Transcriptional activation with fusions of the GAL4 DNA binding domain to N- and C-terminal regions of HOXD-4. GAL4 is stippled. HOXD-4 regions (open rectangles) are lettered as in Figure 4. Numbers below indicate amino acid residues as in Figure 1. C' is a subregion of C as indicated. HA-4 is the HOXA-4 C-terminus. 'FOLD ACTIVATION' gives the increase in luciferase activity relative to that obtained with the GAL4 DNA binding domain (GAL/). Standard deviations and the numbers of independent experiments are shown. '+' under 'WESTERN' indicates that the protein could be readily detected in extracts from transient transfections using polyclonal antibodies directed against the HOXD-4 N-terminus. n.d. not determined.

A transcriptional activation domain in the HOXD-4 N-terminus

While the intact yeast GAL4 protein is a strong transcriptional activator of responsive promoters in eukaryotic cells, the isolated GAL4 DNA-binding domain is not. Fusion of the GAL4 DNA-binding domain to heterologous polypeptides thus provides an assay for transcriptional activation domains distinct from those involved in DNA binding (36). HOXD-4 N-terminal regions A through D were fused to the GAL4 DNA-binding domain and tested for their ability to activate transcription from a GAL4-responsive reporter in differentiating P19 cells. As shown in Figure 5, this fusion product (GAL/ABCD) was 250 times more active than GAL/ which has the GAL4 DNA-binding domain alone (507 fold *vs* 2 fold). This result maps one or more transcriptional activation domains to the proline-rich HOXD-4 N-terminus.

Table 3. Transcription factor AP2 inhibits transcriptional activation by proteins having the HOXD-4 proline-rich N-terminus[a]

Activator Promoter Squelching vector:	HOXD-4 Hoxd-4	HOXD-4 p34a	GAL/ABCD pXPAL7	GAL/C' pXPAL7	GAL/VP16 pXPAL7
none	11 (100)	8 (100)	193 (100)	9 (100)	2,850 (100)
AP2 0.4µg	3 (28)	3 (38)	-	-	-
1.0µg	2 (15)	2 (25)	5 (3)	1 (12)	1,950 (68)
GAL/VP16 1.0µg	26 (229)	-			

[a]Acidic and proline-rich transcrpitional activators were tested for their ability to squelch activation by HOXD-4 and by three GAL4 chimeras. Combinations of activating proteins and luciferase reporters are given on top. pSXΔP2 carries the *Hoxd-4* promoter (Fig. 2B). p34a has multiple HOXD-4 binding sites upstream of the adenovirus major late promoter. pXPAL7 has 5 palindromic binding sites for GAL4 placed upstream of the adenovirus major late promoter (Fig. 5). 0.4 or 1.0 µg of the AP2 expression vector, or 1.0 µg of the GAL/VP16 expression vector was cotransfected with fixed amounts of the HOXD-4 expression vector p4.2 (1 µg/plate) or expression vectors for activating GAL4 fusion proteins (0.1 µg/plate). values are the fold activation over basal promoter levels mediated by activating proteins in the absence and presence of factors tested for squelching. Numbers in brackets are the percentage of activation in the absence of factors tested for squelching.

Deletion of regions A and B from HOXD-4 had no effect on transcriptional activation (Fig. 4), indicating a role for regions C or D in N-terminal function. Fusion of the GAL4 DNA-binding domain to just a portion of the proline-rich region C (GAL/C'; residues 24 to 109) also increased activation GAL/ (Fig.5). This suggests that region C spans a transcriptional activation domain of the proline-rich variety (37).

Deletion and substitution mutants suggested a role in transcriptional activation for the HOXD-4 C-terminus and the equivalent region of the *Hoxa-4* product (Fig. 4). However, fusion of the GAL4 DNA-binding domain to the HOXD-4 C-terminus (GAL/H) had no effect on transcriptional activation (Fig. 5). A similar result was obtained with the HOXA-4 C-terminus (GAL/HA-4). Coexpression of either of these products with GAL/ABCD did not result in any synergistic activation of reporter gene expression (not shown). The role of these C-terminal sequences thus remains to be determined.

The transcription factor AP2 has been shown to harbor a transcriptional activation domain of the proline-rich variety (22) and would be expected to squelch the activity of HOXD-4, if HOXD-4 also contains a proline-rich activation domain. Cotransfections with an expression vector for the AP2 transcription factor revealed an inhibition of HOXD-4 activity on both artificial and *Hoxd-4* promoters (Table 3). By contrast, a GAL/VP16 expression vector did not squelch the activity of HOXD-4 on the *Hoxd-4* promoter (Table 3). In fact, we note a two fold increase in luciferase activity in this case. This is not due to an effect of GAL/VP16 on the *Hoxd-4* promoter (data not shown). Cotransfections with a luciferase reporter under the control of the *pgk-1* promoter showed a 10 and 50% repression by 0.4 and 1.0µg of AP2 expression vector, respectively. It is thus possible that AP2 may be decreasing the production of HOXD-4 from the *pgk-1*-based expression vector. However, this could not account for the repressive effects of AP2 on HOXD-4 activity since decreasing the amount of the HOXD-4 expression vector by 2 fold does not decrease activation of the *Hoxd-4* promoter (data not shown).

Activation of the GAL4-responsive reporter by GAL/ABCD and GAL/C' was also inhibited by cotransfection with the AP2

expression vector (Table 3). AP2 did not significantly squelch the activity of the acidic activator GAL/VP16. Because all GAL4 constructs were expressed from the same promoter, these differences cannot be accounted for by differential effects of AP2 on the expression of *Gal4*-containing genes. These results suggest that AP2 is able to squelch HOXD-4 activity due to competition for mutually required auxiliary factors necessary for the function of proline-rich, but not acidic, transcriptional activation domains.

DISCUSSION

HOXD-4 activates transcription from an autoregulatory element in *Hoxd-4* upstream sequences in transient transfections of P19 cells (1). We used this activity to assess the effects of various deletion and substitution mutations of the *Hoxd-4* coding region. Fusions to the GAL4 DNA-binding domain were also used to assay regions of the HOXD-4 protein for transcriptional activation domains.

The importance of the DNA-binding homeodomain for HOXD-4 activity was clearly established by a single residue change within the recognition helix; conversion of glutamine at position 201 to lysine, a change expected to alter binding-site specificity (34, 35), abrogated transcriptional activation by HOXD-4. We have confirmed that this mutation impairs recognition of the HOXD-4 binding site in band shift assays (not shown). We conclude that the activation of the *Hoxd-4* promoter by HOXD-4 requires the homeodomain, unlike some homeodomain-independent activities noted for other proteins (38, 39).

A pentapeptide motif (consensus VYPWM) (8, 9) is located N-terminal to the homeodomain in Antp class homeoproteins and in the labial subfamily. It is not present in Abd-B, nor in Abd-B-like vertebrate proteins described to date (40). In our assay we did not observe a dramatic change in activity upon deletion of the pentapeptide motif (Fig. 4). Though a possible explanation is that P19 EC cells lack cofactors which could modulate homeoprotein activity through the pentapeptide motif, we obtained similar results in transfections of differentiated P19 cells (Fig. 4). We were also unable to detect activity for conserved regions A and B at the extreme N-terminus of HOXD-4. Conservation suggests function, therefore we conclude that transient transfections of P19 cells cannot reveal all the biochemical activities of homeoprotein molecules.

C-termini have been implicated in quantitative roles in homeoprotein function (41). Deletion of residues C-terminal to the HOXD-4 homeodomain (dlGH, Fig. 4) reduced activity on the *Hoxd-4* promoter, and this could be rescued by replacement with the HOXA-4 C-terminus. However, neither was able to confer activity on the GAL4 DNA-binding domain (GAL/H and GAL/HA-4, Fig. 5). Nor did they cooperate with the HOXD-4 N-terminus in cotransfections with GAL/ABCD (not shown). Others report a mild (4 fold) increase in HOXD-4 activity upon deletion of regions G and H (42). However, these earlier studies made use of high levels of transfected *Hoxd-4* expression vectors (5µg/plate) in HeLa cells. We have also noted 4 to 6 fold increases over wild-type in the activity of HOXD-4 C-terminal deletions at ≥4µg/plate of transfected expression vector (unpublished observations). The significance of these increases at such high vector concentrations is not clear.

In Dfd, the region N-terminal to the homeodomain is 365 residues long, whereas in HOXD-4 it is 149 residues. Within

these regions, there is little sequence conservation apart from the extreme amino terminus and the region of the pentapeptide. Moreover, we have no direct evidence for transcriptional activation domains in either of these conserved regions. N-terminal to the homeodomain, Dfd has regions up to 10% in proline, but has domains richer in histidine, serine, glycine and acidic residues (43). By contrast, the HOXD-4 N-terminal half is proline-rich and displayed strong transcriptional activation function when fused to the GAL4 DNA-binding domain (GAL/ABCD, Fig. 5). Region C of HOXD-4 is 21% in proline. A 74 residue stretch of region C (residues 48 to 121) is 30% proline. Most of region C was fused to the GAL4 DNA-binding domain and conferred an 11 fold activation above background (GAL/C', Fig. 5). This is less than with the whole HOXD-4 N-terminus (371 fold), but comparable to that seen with the CTF/NF-1 proline-rich region fused to the SP1 DNA-binding domain (5 fold)(44).

The decreased activity of GAL/C' suggests additional activity in regions A, B or D. However, deletion of these regions did not impair HOXD-4 function (Fig. 4). It should be noted that 7 proline residues between positions 109 and 141 are missing from GAL/C', and the absence of this stretch may have been detrimental. Moreover, using extracts of transiently transfected cells, GAL/C' gave a less intense retarded band than GAL/ABCD in electrophoretic mobility shift assays with a GAL4 binding site (not shown). It is thus not clear for the moment whether the reduced activity of GAL/C' relative to GAL/ABCD is due to the loss of complementing residues. A HOXD-4 protein containing a duplication of residues 6 to 81 appears to be much more active than the wild-type protein (42). This stretch is proline- and glycine-rich and straddles much of the region we have called C', supporting a role for these residues in transcriptional activation.

Squelching (45) is a phenomenon whereby one transcription factor can inhibit the activity of another independent of both its ability to bind DNA and to directly interact with the inhibited protein. AP2 has a proline-rich transcriptional activation domain (22) and is able to squelch HOXD-4 activity, but not that of the acidic transcriptional activator GAL/VP16 (Table 3). Neither is GAL/VP16 able to squelch HOXD-4. These data support the presence of functionally similar proline-rich transcriptional activation domains in AP2 and HOXD-4 that are distinct from that of GAL/VP16. AP2 expression overlaps with that of many Hox genes both spatially and temporally in development. Tissues expressing both include neural crest, limb bud mesenchyme and the meso- and metanephros (46). Overlapping expression and similar proline-rich transcriptional activation domains may suggest cooperation between HOX proteins and AP2 in embryonic patterning.

Our results strongly suggest that a transcriptional activation domain of the proline-rich type (37) contributes to the activity of the HOXD-4 N-terminus, and that proline-rich domains of other homeoproteins may function in a similar capacity. HOX proteins with proline-rich N-terminal regions ($\geq 20\%$ over at least 60 residues) include the HOXD-4 paralogs HOXA-4 (11), HOXB-4 (12) and HOXC-4 (47). Others have suggested that the proline-rich N-terminus of chicken HOXA-4 functions as a transcriptional activator, though their results are difficult to interpret due to competitive effects on the high endogenous activity of a Jun-responsive reporter (48). By contrast, the background in our GAL-4 assay is low.

Dfd itself does not have proline-rich domains. Nonetheless, human HOXD4 can substitute for Dfd for the activation of the *Dfd* transcription unit in flies (49). In addition, the *Dfd* autoregulatory enhancer may respond to Dfd subfamily Hox products in mice (50). These data suggest that for some of the genetic pathways regulated by Dfd subfamily members, transcriptional activation can be accomplished by disparate domains. We also note that many *Hox* genes do not encode proline-rich regions, suggesting that these proteins use functionally distinct transcriptional activation domains. In the embryo, where the products of multiple *Hox* genes may bind responsive enhancers in various combinations, the specific array of such regulatory domains could be an important determinant of the transcriptional response.

ACKNOWLEDGEMENTS

REFERENCES

1. Pöpperl, H. and Featherstone, M.S. (1992) *EMBO J,* **11**, 3673–3680.
2. McGinnis, W. and Krumlauf, R. (1992) *Cell,* **68**, 283–302.
3. Kuziora, M.A. and McGinnis, W. (1988) *Cell,* **55**, 477–485.
4. Regulski, M., Dessain, S., McGinnis, N. and McGinnis, W. (1991) *Genes Dev,* **5**, 278–286.
5. Laughon, A. and Scott, M.P. (1984) *Nature,* **310**, 25–31.
6. McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A. and Gehring, W.J. (1984) *Cell,* **37**, 403–408.
7. Scott, M.P. and Weiner, A.J. (1984) *Proc. Natl. Acad. Sci. USA,* **81**, 4115–4119.
8. Boncinelli, E., Simeone, A., LaVolpe, A., Faiella, A., Fidanza, V., Acampora, D. and Scotto, L. (1985) *Cold Spring Harbor Symp. Quant. Biol.,* **50**, 301–306.
9. Mavilio, F., Simeone, A., Giampaolo, A., Faiella, A., Zappavigna, V., Acampora, D., Poiana, G., Russo, G., Peschle, C. and Boncinelli, E. (1986) *Nature,* **324**, 664–668.
10. Featherstone, M.S., Baron, A., Gaunt, S.J., Mattei, M.-G. and Duboule, D. (1988) *Proc. Natl. Acad. Sci. USA,* **85**, 4760–4764.
11. Galliot, B., Dollé, P., Vigneron, M., Featherstone, M.S., Baron, A. and Duboule, D. (1989) *Development,* **107**, 343–359.
12. Graham, A., Papalopulu, N., Lorimer, J., McVey, J.-H., Tuddenham, E.G.J. and Krumlauf, R. (1988) *Genes Dev,* **2**, 1429–1438.
13. Scott, M.P. (1992) *Cell,* **71**, 551–553.
14. Adra, C.N., Boer, P.H. and McBurney, M.W. (1987) *Gene,* **60**, 65–74.
15. Kozak, M. (1987) *Nucleic Acids Res,* **15**, 8125–8148.
16. Sasaki, H., Yokoyama, E. and Kuroiwa, A. (1990) *Nucleic Acids Res,* **18**, 1739–1747.
17. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Methods Enzymol,* **154**, 367–382.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463–5467.
19. Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E. and Chambon, P. (1989) *Cell,* **59**, 477–487.
20. White, J., Brou, C., Wu, J., Lutz, Y., Moncollin, V. and Chambon, P. (1992) *EMBO J,* **11**, 2229–2240.
21. Nordeen, S.K. (1988) *BioTechniques,* **6**, 454–457.
22. Williams, T. and Tjian, R. (1991) *Genes Dev.,* **5**, 670–682.
23. Spindler, K.R., Rosser, D.S.E. and Berk, A.J. (1984) *J. Virol.,* **49**, 132–141.

24. Harlow, E. and Lane, D. (1988) Antibodies: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor.
25. Kissinger, S.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) *Cell, 63,* 579–590.
26. Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991) *Cell, 67,* 517–528.
27. Otting, G., Qian, Y.Q., Müller, M., Affolter, M., Gehring, W. and Wüthrich, K. (1988) *EMBO J, 7,* 4305–4309.
28. Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J. and Wüthrich, K. (1989) *Cell, 59,* 573–580.
29. Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) *EMBO J, 9,* 3085–3092.
30. Percival-Smith, A., Müller, M., Affolter, M. and Gehring, W.J. (1990) *EMBO J, 9,* 3967–3974.
31. Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. and Gehring, W.J. (1990) *Proc. Natl. Acad. Sci. USA, 87,* 4093–4097.
32. Müller, M., M. Affolter, M., W. Leupin, W., G. Otting, G., K. Wüthrich, K. and Gehring, W.J. (1988) *EMBO J, 7,* 4299–4304.
33. Hanes, S.D. and Brent, R. (1991) *Science, 251,* 426–430.
34. Hanes, S.D. and Brent, R. (1989) *Cell, 57,* 1275–1283.
35. Treisman, J., Gönczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989) *Cell, 59,* 553–562.
36. Ma, J. and Ptashne, M. (1987) *Cell, 51,* 113–119.
37. Mitchell, P.J. and Tjian, R. (1989) *Science, 245,* 371–378.
38. Zappavigna, V., Renucci, A., Izpisa-Belmonte, J.-C., Urier, G., Peschle, C. and Duboule, D. (1991) *EMBO J., 10,* 4177–4187.
39. Fitzpatrick, V.D., Percival-Smith, A., Ingles, C.J. and Krause, H.M. (1992) *Nature, 356,* 610–612.
40. Erselius, J.R., Goulding, M.D. and Gruss, P. (1990) *Development, 110,* 629–642.
41. Gibson, G., Schier, A., LeMotte, P. and Gehring, W.J. (1990) *Cell, 62,* 1087–1103.
42. Lufkin, T., Mark, M., Hart, C.P., Dollé, P., LeMeur, M. and Chambon, P. (1992) *Nature, 359,* 835–841.
43. Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987) *EMBO J, 6,* 767–777.
44. Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) *Cell, 58,* 741–753.
45. Gill, G. and Ptashne, M. (1988) *Nature, 334,* 721–724.
46. Mitchell, P.J., Timmons, P.M., Hébert, J.M., Rigby, P.W.J. and Tjian, R. (1991) *Genes Dev., 5,* 105–119.
47. Geada, A.M., Gaunt, S.J., Azzawi, M., Shimeld, S.M., Pearce, J. and Sharpe, P.T. (1992) *Development, 116,* 497–506.
48. Sasaki, H., Yamamoto, M. and Kuroiwa, A. (1992) *Mechanisms Dev, 37,* 25–36.
49. McGinnis, N., Kuziora, M.A. and McGinnis, W. (1990) *Cell, 63,* 969–976.
50. Awgulewitsch, A. and Jacobs, D. (1992) *Nature, 358,* 341–344.
51. De Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. and Subramani, S. (1987) *Mol. Cell. Biol., 7,* 725–737.
52. Cianetti, L., Di Cristofaro, A., Zappavigna, V., Bottero, L., Boccoli, G., Testa, U., Russo, G., Boncinelli, E. and Peschle, C. (1990) *Nucleic Acids Res, 18,* 4361–4368.
53. Pöpperl, H. and Featherstone, M.S. (1993) *Mol. Cell. Biol., 13,* 257–265.